

Starter Questions

Out of the following types of data, decide which is continuous and which is discrete:

The lengths of some roads.

The number of 'cats eyes' on a one mile stretch of road.

The time it takes twenty students to complete an English essay.

The number of pages in twenty students English essays.

The weights of sacks of potatoes.

The number of potatoes in some sacks of potatoes.

The depth of water as the tide comes in and goes out.

9.1 Sampling

Everything in statistics focuses on data.

9.1 Sampling

Populations

In any statistical investigation there will be a group of something (it could be people, items, animals... or anything else) that you want to find out about.

The whole group consisting of every single person/item is called the **population**.

A **parameter** is a number that describes the entire population, e.g. the mean of a population.

Parameters for a population can be almost

9.1 Sampling

Finite Populations

It is possible for someone to count how many members there are:

- Fish in an aquarium
- Trees in a garden
- To collect information about a population can do a **survey** (we question people or examine the items to find out the required data)

Infinite Populations

It is impossible to know how many members there are (too many to count!):

- Fish in the Atlantic ocean
- Leaves in a forest
- Stars in the sky
- People in the world

9.1 Sampling

Censuses

A **census** is a survey of the **whole** population.

Advantage

- It's an accurate representation of the population because every member has been surveyed (it's **unbiased**)

9.1 Sampling

Censuses

A **census** is a survey of the **whole** population.

Disadvantages

- For large populations it takes a lot of time and effort to carry out
- It can be very expensive
- It is difficult to make sure all members of the population are surveyed
- If the tested items are used up or damaged, a census can be impractical

Tip: watch out for anything that might make doing a census a silly idea

9.1 Sampling

Sampling

A **sample** is a subset of data from the population.

Questioning or examining a sample from a population is often more practical (much quicker and easier).

Data collected from a sample is used to draw conclusions about the whole population so it must be a **representative sample**

A **statistic** is a number taken from the sample that can be used to estimate the related

9.1 Sampling

Sampling

A **sample** is a subset of data from the population.

Advantages

- Sampling surveys are quicker and cheaper than a census
- It's the only option when survey items are used up or damaged

9.1 Sampling

Sampling

A **sample** is a subset of data from the population.

Disadvantages

- There is variability between samples, each possible sample may give different results, so you may select one sample that doesn't accurately represent the population
- They can easily be affected by **sampling bias**

9.1 Sampling

Sampling

A **sample** is a subset of data from the population.

To avoid sampling bias

- Select from the correct population and make sure no member is **excluded**
- Select your sample at **random**
- Make sure the sampling members **respond**

9.1 Sampling

Example 1

A student wants to know the mean number of sweets in a packet of their favourite snack. They open 10 packets and count the number of sweets in each. They find the mean of these totals.

Identify the population, the parameter, the sample and the statistic in this example, and say which statistic can be used to estimate the parameter.

9.1 Sampling

Example 1 - Answer

Population – all packets of the snack

Parameter – the mean number of sweets per packet

Sample – the 10 packets of sweets

Statistic – the mean number of sweets in the 10 packets

The sample mean is an estimator for the parameter

Optional: Ex1.1 from CGP